BIOLOGICAL RESEARCH METHOD

★

A Practical Guide

H. H. Holman

BIOLOGICAL RESEARCH METHOD

★

A Practical Guide

H. H. HOLMAN

D.Sc., Ph.D., F.R.C.V.S.

Deputy Director, Agricultural
Research Council Field Station
Compton, Berks

OLIVER & BOYD

EDINBURGH AND LONDON

OLIVER AND BOYD LTD
Tweeddale Court
Edinburgh

39a Welbeck Street
London W.1

First published 1962

Printed in Great Britain by
Oliver & Boyd, Tweeddale Court, Edinburgh

PREFACE

This book is intended as a personal guide for research workers who carry out experiments involving animals or their products.

Without guidance a scientist, although giving full rein to his imagination, may neglect the main factor that will build his reputation. This factor is reliability; without this his work will never be spoken of with respect. It is important to acquire reliability early, for a few years spent on experiments that lack adequate controls will lead to an ill-disciplined technique which is hard to remedy. This book offers guidance on the design, recording and analysis of experiments according to rules used in realistic research.

One of the tools of the trained research worker is biometrics because, even if his data are analysed by a professional statistician, he still needs to know the implications that can be drawn from these measurements, how his data should be displayed, and the logic of experimental design. This book introduces these subjects and it also offers the reader a selection of simple tests from which he may choose those that suit his purpose.

ACKNOWLEDGEMENTS

I would like to express my thanks to I. H. Pattison, B.SC., M.R.C.V.S., and to G. B. S. Heath, B.SC., F.R.C.V.S., for their criticism of the first draft and for their very kind encouragement. My task was made easier by having worked and exchanged ideas with Pattison over many years and I am also obliged to him for having read the proofs.

In trying to understand and explain the basic terms and rationale of statistical methods I am very indebted to my brother, Lt.-Col. L. J. Holman, B.SC., F.B.PS.S., for answering, patiently and with lucidity, many questions over a large number of years.

The abridged Tables III, IV and V are from the statistical tables of Fisher and Yates detailed in the bibliography, and I am grateful for the authors' permission to publish them. I also wish to thank Prof. D. Mainland for his permission to use part of his Table III from the binomial tables detailed in the references and in the bibliography.

There is little that is original in the text except, perhaps, in the undetected errors, and I would be most grateful if readers would point these out to me.

H. H. HOLMAN

CONTENTS

NOTES

Tables and Figures take their numbers from the pages on which they appear. Specific references are detailed at the end of each chapter. General references are to books in the Selected Bibliography

CHAPTER 1

TRUTH, LOGIC AND CHANCE

> **Man knows that, at any moment, he can tell a lie that, for a while will delay or divert the working of cause and effect. Being an animal who is still learning to reason, he does not yet understand why, with a little more, or a little louder, lying, he should not be able permanently to break the chain of that law.**
>
> RUDYARD KIPLING

A biologist engaged in research work should be able to carry out a workmanlike experiment. A workmanlike experiment is one that is well planned, well carried out, and well recorded. As a result the experimental data are easy to assess and the experiment is easy to write up. Such a technique is best learnt from personal supervision but this book attempts to show the beginner some of the basic principles needed to produce reliable and definite results. It should also help the beginner to avoid those peculiar indefinite results which still appear in some journals and which read something like this: "The experiment does not permit us to say definitely that the addition of this new substance to the food always produced an advantageous effect, but animals number 7 and number 11 appeared to show great improvement; and although there was some loss in number 3 this perhaps could be accounted for by assuming that this animal was much older than the others."

It is also hoped that it will encourage readers to avoid results which, if truthfully recorded, might read like this: "It appears likely that there was some rhythm present in these measurements, but it was not possible to prove this because the writer always liked to leave the laboratory at 5 o'clock and made no measurements on Saturdays and Sundays."

TRUTH

Scientific experiments are carried out on objects and animals in order to establish the truth. Most people believe it is quite easy to establish the truth but the old adage states that "Truth lies at the bottom of the well." Things lying at the bottom of a well are difficult to get up, in fact when a dead cat was reported to be lying at the bottom of one of the local wells, all the authorities did was to put up a notice saying that the water was unfit for human consumption. Whether truth is fit for human consumption is a debatable point, but certainly there are more people concerned in its suppression than there are in its production. In trying to proclaim the truth, science is often in conflict not only with politics, religions and commerce but with individual scientists.

However, it is still the grasping of the truth that is the most difficult task, and the unfortunate thing is that this appears to be an easy thing to do and we all feel that we are particularly good at it. Yet the fact remains that when senior research workers are told that recent work has shed new light on a subject, the question immediately asked is, not what method was used, or where it was done, but who did it? The work of one man will be accepted immediately; the work of another will not be, however low the probability of his statistical analysis. The fact that after years of doing research work some workers can still produce results that are suspect is sufficiently interesting to warrant a little closer look at truth.

The first thing to recognise is that truth is not an instinct. On the contrary, in all but the highest civilisations, man's instincts for preservation are served better by his ability to deceive. In trying to explain the nature of truth Trotter (1916), postulated that with gregarious animals there was a "herd instinct" which made the individual instinctively suggestible to the wishes of the herd. In man this instinct provided such emotions as patriotism, religion, esprit de corps and racial and class prejudices. The suggestibility of this instinct was in some cases so powerful that it could overcome even the individual instinct of self-preservation.

Thus, where our experimental findings appear to support the belief or opinion of our group, then conscience is not a reliable guide to the truth, and we tend to accept the result without question, in fact, to quote Trotter:

"When, therefore, we find ourselves entertaining an opinion about the basis of which there is a quality of feeling which tells us that to inquire into it would be absurd, obviously unnecessary, unprofitable, undesirable, bad form, or wicked, we may know that that opinion is a non-rational one, and probably, therefore, founded upon inadequate evidence."

If this is the case, how can we ever get to the state where we wish to know the truth for its own sake? Trotter suggests that "The solution would seem to lie in seeing to it that suggestion always acts on the side of reason; if rationality were once to become really respectable, if we feared the entertaining of an unverifiable opinion with the warmth with which we fear using the wrong implement at the dinner table, if the thought of holding a prejudice disgusted us as does a foul disease, then the dangers of man's suggestibility would be turned into advantages. We have seen that suggestion already has begun to act on the side of reason in some small part of the life of the student of science, and it is possible that a highly prophetic imagination might detect here a germ of future changes."

This suggestibility must be sought by acquiring a liberal education and by mixing with other truth seekers in universities or as a member of scientific societies. It is only when the scientific method of "truth by verification" has become a matter of conscience that a research worker can regularly design the worth while experiment which is intended, not to bolster up an opinion that he feels will bring him fame because of its value to the community, but to test it in such a way that, where necessary, it will show his opinion to be completely untrue.

This attitude of mind is the most important factor in good research work, but it is not sufficient in itself, for good experimentation is both a discipline and an art. And even men who, through education and shrewd observation, have attained eminence in a biological science, can produce pitiably inadequate experiments if they lack training in experimental method.

Art cannot be taught, it can only be demonstrated. It is therefore fortunate that the greater the art of the experimenter the more simple and direct will be his experiments, so that in acquiring the simple methods of logic and biometrics mentioned in this book the novice is learning to use the tools of the expert.

LOGIC

Second to the desire for truth comes the need for a reliable method of reasoning. The process of reasoning that we are accustomed to use in our daily life is that referred to as "Post hoc ergo propter hoc."—after that therefore because of that—and this method is based on the fact that the cause must precede the effect. Thus if a man and a dog are nearly dying from thirst and they reach a stagnant pond, the man may hang back and let the dog drink the water first. If after drinking the dog runs round in circles and then falls over dead, it is commonsense for the man to accept the "post hoc" argument that the water is poisonous; if the man is an anti-vivisectionist he will keep the dog back and drink the water himself first, it is then up to the dog to apply a "post hoc" argument.

Although the "post hoc" argument plays a large and useful part in everyday affairs, it can be very misleading. The man who, after a night out in which has he mixed every sort of drink, comes home feeling sick and tells his wife that it must have been something he ate, has probably chosen the wrong antecedent as the cause, whereas the mother who explains that her little girl has lovely teeth, curly hair, and never has a cold, because she has always been fed on wholemeal bread, may not have selected the right attributes as effects.

In experiments on animals, in addition to these wrong conclusions based on faith or self-deception, mistaken interpretation may be due to the natural waxing and waning of physiological or disease processes, so that the "post hoc" argument becomes quite useless as a weapon in the search for truth. Nevertheless it must be admitted that in spite of this there are committees that even today, permit observations of this type in which the only thing that is certain is that the result will be debatable. The use of such an inefficient method of inference in experiments involving domesticated animals is often encouraged by the owner, who argues that if the diet, drug, or vaccine is going to benefit his animals, then all of them should have it. With field trials that involve people, the emotional factor is paramount and Sinclair Lewis, in his book *Martin Arrowsmith*, recorded very clearly the conflict between a research worker fighting for the truth, and his companions, moved by pity to take any step, however illogical, to reduce the suffering caused by an outbreak of plague.

In experimental work the "post hoc" argument is avoided wherever possible and it is usually replaced by the "method of difference." This method lays down that if you have two sets of circumstances that are alike in every respect except one, and if the phenomenon under investigation occurs in the group containing this one circumstance, then this circumstance is the cause, or the effect, or is part of the cause or effect of the phenomenon under investigation.

Assume that we were going to use 20 mice in a simple vaccination experiment; that all the mice were alike, and that they would be kept in the same environment. We would split the mice into two groups of 10 and give the vaccine to one group. Later we would give a test dose to all the mice. If the ten vaccinated mice survived and the ten unvaccinated, or control, mice died, then we could say that the single circumstance that was different—the injection of a vaccine—was the cause or effect, or part of the cause or effect, of the phenomenon under investigation, which in this instance was the survival of the mice. Further, we could say that as the vaccine was given before the phenomenon occurred it was the cause, and not the effect, of the phenomenon.

Statistical method

Unfortunately with biological material the "method of difference" is often an ideal that we cannot reach. The mice may all look alike but some, because of their physiological make-up will be more susceptible or less susceptible than others, while in addition some may have suffered from mild and undetected diseases that have increased, or decreased, their natural susceptibility. Thus when a test dose is given, instead of having 10 deaths in one group and 10 survivors in the other, we may find that we have both dead mice and live ones in each group. It is with this type of result that the trained research worker uses a statistical test to provide an objective assessment of the result.

Because of these hidden factors we can regard mice as being like playing cards, with their outward appearance as similar to the backs of a pack of cards and their intrinsic values as varying as the unseen faces of the cards. Thus a card might belong to a black suit or a red one, it might bear a high number or a low one, and the number might be divisible by 3 or 5. Each of these possibilities might represent some crucial factor in the game to be played.

If, using cards, our experiment was to find which was the better of two players, we could ask them to play but our judgement would be tempered by the fact that the winner, instead of being the better player, might merely have held the better cards. All we could do to prevent this happening would be to make sure that the cards were well shuffled before they were dealt.

Statistical tests are based on the knowledge obtained from such actions as selecting playing cards, tossing coins or withdrawing different coloured balls from a bag. The results from such frivolous material have proved so useful that they are widely used in industry where false assumptions are punished by financial loss. A fundamental condition of this method is that before any card is selected the pack must be well shuffled; similarly the balls in the bag must be well mixed or the coin tossed in a fair manner.

If, therefore, we are going to regard the mice as similar to playing cards, and interpret the result by statistical methods, the theorist can maintain that the mice must be distributed into the two groups by some method by which they are shuffled, mixed or tossed; this process is known as randomisation. Methods of tossing or mixing mice might add excitement to laboratory life, but the same result is obtained by catching mice in sequence and deciding their group by tossing a coin or using some other method of chance.

With the contents of each group chosen by chance we apply our experimental interference to one group—the experimental group—and leave the other group—the control group—untreated. To judge the result we adopt the line of reason that, as circumstances vary in each group, we cannot use the method of difference. We can, however, say that we distributed the unknown circumstances as equally as possible into the two groups by using a method of chance. Therefore if, when the experiment is ended, one group varies from the other to a degree that would be unlikely to be due to chance alone, then there is an objective reason to accept that the treatment had produced some effect.

Thus the result has not been proved, if we had to do that the advance of science would be very slow, but the hypothesis has been justified, and Arber (1954) writing on the philosophy of science, accepts the ruling that all that is necessary for a forward step is that an hypothesis should be justified.

Selection into groups

To make the experimental and the control groups comparable there are two methods of selection. First, with measurable qualities we can try and keep the two groups as equal as possible and, second, where qualities are unseen and unmeasurable, we can distribute them by chance, making sure that each mouse has an equal chance of falling into the experimental or control group.

It is, of course, no good taking into account measurable qualities that experience has shown will not affect the result. For example, although all the mice may look alike, accurate weighing would show that each mouse was of a different weight, but unless we believed that this would affect the result, this variation would be ignored. The only measurable factor worth taking into account might be the effort made by each mouse to evade capture, and if it was felt that this alertness was an indication of good health, the following method of selection could be used. Prepare a cage for each of the two groups. Catch two mice and toss a coin to decide which cage the first mouse should be put into; putting the second mouse into the other one. Continue this sequence until all the mice are allocated and then decide by a final toss which lot should act as the experimental group.

Analysis

After allocating the 20 mice into experimental and control groups, the experimental mice will receive the vaccine and, after this has had time to become effective, the test dose will be given to all the mice. At an appropriate time after this the survivors in both groups will be counted.

Turning now to the analysis of the vaccination experiment, let us assume that in the result two vaccinated and six control mice have died. This type of result is often put into what is called a " two by two " or a four-cell table, as follows:

	Died	Survived	Sub-totals
Vaccinated	2	8	10
Controls	6	4	10
	8	12	20

A worker looking at these results might say " As I arranged these two groups to be comparable, my results show that, as I expected, the vaccine has given some protection." If asked about the effect of chance, and if at the stage where he still transferred everything to percentages, he might present his results thus "100 per cent more mice have survived in the vaccinated group compared with the control group, or taking it the other way round, 200 per cent more mice have died in the control group than in the vaccinated group, which to me appears significant." With similar figures and a strong desire to be right, many long arguments can ensue if the worker is once put on the defensive.

To avoid these arguments, statisticians have laid down an objective standard as a guide as to whether an experiment has justified the hypothesis or not. This standard lays down that an experiment has not justified the hypothesis unless the figures in the result could not occur by chance alone more than once in twenty trials.

This criterion is based on a "Null Hypothesis" by which it is assumed that the experimental interference of vaccination has had no effect whatsoever. The deathrate in each group is therefore a matter of luck, dependent on how the naturally resistant mice were selected into each group. If, with 8 susceptible mice and 12 resistant mice, the number of survivors in each group could easily occur by chance alone then the null hypothesis is sustained, and there is no reason to alter the assumption that the vaccine has played no part in the result. If, however, the figures in the result could occur only once in 20 trials the result is accepted as sufficiently unusual to justify the assumption that some other factor has played a part. Therefore, if—and only if—the two groups have been treated in such a way that the only real difference in treatment between them was the vaccination of one group, then it is assumed that this vaccination has influenced the result.

Trial and error method

How can we find the likelihood of the various ways in which the 8 susceptible mice fall into the two groups of 10? If we obtain 20 rubber stoppers of the same size and mark 8 of them to represent the susceptible mice, we can use these symbols to repeat our experiment as many times as we like. The stoppers are placed in a bag which is well shaken to mix them up and the withdrawals are made blindfold. To imitate the first method we used with mice, one stopper would be withdrawn from the bag and a coin tossed to see if it was in the " heads " or " tails " group and then a second stopper would be withdrawn and placed in the other group. This procedure would be continued until all the stoppers had been dealt with, when the result would be recorded.

With a little experience it would be realised that this close imitation of the experimental method was unnecessary and that similar results could be obtained by picking out unseen stoppers several at a time, until 10 had been withdrawn. To build up what is called a " Frequency Table " we would record the number of marked stoppers out of the 10 chosen. First we would draw up a table with columns headed 1, 2, 3, . . . up to 8. We would then mix the stoppers in a box and withdraw 10 blindfold. We would then count the number of marked stoppers present, and if there were four we would make a mark under the 4 column. Finally, we would count the number of strokes under each column and these totals would give a frequency table. Using this method 100 trials were carried out with the following result:

Number of Marked Stoppers Withdrawn	0	1	2	3	4	5	6	7	8	Total Trials
Frequency with which they were withdrawn	0	0	9	30	28	26	7	0	0	100

Having obtained this frequency table we can, if we wish, fill in the other events. Thus the result of withdrawing no marked stoppers signifies that there were no deaths in the first group of 10 and that here there must be 10 survivors. As there were no deaths in the first group then the second group must contain all the 8 deaths and only 2 survivors. Working on the same lines for the other possibilities we can build up the following table:

Selected Group	Deaths	0	1	2	3	4	5	6	7	8
	Survivals	10	9	8	7	6	5	4	3	2
Remaining Group	Deaths	8	7	6	5	4	3	2	1	0
	Survivals	2	3	4	5	6	7	8	9	10
Frequency		0	0	9	30	28	26	7	0	0

From this table it can be seen that our result of 2 deaths in one group, which necessitates 6 falling in the other group, has occurred 9 times with the 2 falling in the selected group and 7 times with the 2 falling in the remaining group, giving a total of 16 times in a 100. Hence instead of occurring only once in 20 times, as set by the criterion, our result would occur about once in every 6 trials; thus our experimental trials have shown that such a result could happen fairly often by chance alone and that there was no need to postulate that the vaccine had produced any effect at all.

Permutations

We have discussed an experiment involving mice and have used the figures suggested for the result as the basis of another experiment in which, by using stoppers as symbols, and repeating the experiment 100 times, we were able to demonstrate that the result could not be considered statistically significant.

Must we always use stoppers, or is there an easier and more reliable arithmetical method of working out the odds ? Yes, there is an easier way which makes use of permutations, so that perhaps it would be best to remind readers what this term means before describing the test. Permutations are the numbers of different ways in which the number of objects you are dealing with can be arranged.

If the number of objects is 5, then you can have any of the 5 as the first object and any of the remaining 4 as the second object, so that each of the first 5 ways has 4 alternatives as second choice, giving 5×4, or 20, ways of arranging the first 2 objects. With each of these 20 ways there are 3 different objects possible for the third position and 2 alternatives for the fourth position with the remaining object for the fifth and last position. The possible ways of arranging them are; $5 \times 4 \times 3 \times 2 \times 1$, which is called "factorial five" and written either as $\underline{5|}$ or $5!$. Thus with the 5 letters in a bag the chance of spelling a word like "brain" correctly by pulling the 5 letters at random from a bag would be

$\frac{1}{5!}$ or 1 in 120.

In some other five-letter words the same letter may occur more than once; for example the word "peels" has 2 letters the same. It seems obvious, and if it is not you can prove it by experiment, that there is the same chance of spelling "peels" correctly as there is of obtaining some other order of letters such as "plsee," and this latter combination is more convenient for explanation.